

# Whole-exome sequencing identifies pathogenic variants in *TJP1* gene associated with Arrhythmogenic Cardiomyopathy

De Bortoli et al. *TJP1* pathogenic variants in ACM

Marzia De Bortoli, PhD<sup>1</sup>; Alex V. Postma, PhD<sup>2</sup>; Giulia Poloni, PhD<sup>1</sup>; Martina Calore, PhD<sup>1</sup>; Giovanni Minervini, PhD<sup>3</sup>; Elisa Mazzotti, MD, PhD<sup>4</sup>; Ilaria Rigato, MD, PhD<sup>4</sup>; Micaela Ebert, MD<sup>5,6</sup>; Alessandra Lorenzon, PhD<sup>1</sup>; Giovanni Vazza, PhD<sup>1</sup>; Alberto Cipriani, MD<sup>4</sup>; Riccardo Bariani, MD<sup>4</sup>; Martina Perazzolo Marra MD, PhD<sup>4</sup>; Gaetano Thiene, MD<sup>4</sup>; Luciano Daliento MD<sup>4</sup>; Domenico Corrado, MD, PhD<sup>4</sup>; Cristina Basso, MD, PhD<sup>4</sup>; Silvio C.E. Tosatto, PhD<sup>3,7</sup>; Barbara Bauce, MD, PhD<sup>4</sup>; J. Peter van Tintelen MD, PhD<sup>8,9</sup>; Alessandra Rampazzo, PhD<sup>1</sup>.

<sup>1</sup>Department of Biology, University of Padua, Padua, Italy. <sup>2</sup>Department of Medical Biology and Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands.

<sup>3</sup>Department of Biomedical Sciences, University of Padua, Padua, Italy. <sup>4</sup>Department of Cardiac, Thoracic and Vascular Sciences, University of Padua, Padua, Italy. <sup>5</sup>Department of Electrophysiology, Heart Center, University of Leipzig, Leipzig, Germany. <sup>6</sup>Department of Cardiology, Leiden University Medical Center, Leiden, The Netherlands. <sup>7</sup>CNR Institute of Neuroscience, Padua, Italy. <sup>8</sup>Department of Clinical Genetics, Amsterdam University Medical Center, University of Amsterdam, Amsterdam, The Netherlands. <sup>9</sup>Department of Genetics,

Corresponding author:

Alessandra Rampazzo

Department of Biology, University of Padua

Via G. Colombo 3, 35131 Padua, Italy

Phone: +39 049 8276208; fax: +39 049 8276209; e-mail: [alessandra.rampazzo@unipd.it](mailto:alessandra.rampazzo@unipd.it)

**Word count:** 5919

**Subject Terms:** Sudden Cardiac Death, Cardiomyopathy, Genetics.

## Abstract

**Background:** Arrhythmogenic Cardiomyopathy (ACM) is an inherited cardiac disease characterized by progressive fibro-fatty myocardial replacement, ventricular arrhythmia, heart failure and sudden death. Causative mutations can be identified in 60% of patients and most of them are found in genes encoding mechanical junction proteins of the intercalated disc (ID).

**Methods:** Whole-exome sequencing (WES) was performed on the proband of an ACM family. Sanger sequencing was used to screen for mutations the tight junction protein 1 (*TJPI*) gene in unrelated patients. Predictions of local structure content and molecular dynamics simulations were performed to investigate the structural impact of the variants.

**Results:** A novel c.2006A>G p.(Y669C) variant in *TJPI* gene was identified by WES in an ACM patient. *TJPI* encodes zonula occludens 1, an ID protein interacting with proteins of gap junctions and area composita. Additional rare *TJPI* variants have been identified in 1 out of 40 Italian probands (c.793C>T p.(R265W)) with arrhythmogenic right ventricular cardiomyopathy and in 2 out of 43 Dutch/German patients (c. 986C>T, p.(S329L) and c.1079A>T, p.(D360V)) with dilated cardiomyopathy and recurrent ventricular tachycardia. The p.(D360V) variant was identified in a proband also carrying the p.(I156N) pathogenic variant in *DSP*. All 4 *TJPI* variants are predicted to be deleterious and affect highly conserved amino acids, either at the guanylate kinase-like (GUK) domain (p.(Y669C)) or at the disordered region of the protein between the PDZ2 and PDZ3 domains (p.(R265W), p.(S329L) and p.(D360V)). The local unfolding induced by the former promotes structural rearrangements of the GUK domain, whereas the others are predicted to impair the function of the disordered region. Furthermore, rare variants in *TJPI* are statistically enriched in ACM patients relative to controls.

**Conclusions:** We provide here the first evidence linking likely pathogenic *TJPI* variants to ACM. Prevalence and pathogenic mechanism of *TJPI*-mediated ACM remains to be determined.

**Key Words:** Arrhythmogenic cardiomyopathy, pathogenic variant, *TJPI*

## Introduction

Arrhythmogenic cardiomyopathy (ACM) (MIM #107970) is the second most common cause of unexpected sudden death among young people and athletes.<sup>1,2</sup> Its main clinical features comprise structural and functional abnormalities of the ventricles and ventricular arrhythmias.<sup>2,3</sup> The subform predominantly involving the right ventricle (ARVC) is the most common, even though left and biventricular forms are also well-documented.<sup>4,5</sup>

ACM's distinct histopathological profile consists of the progressive replacement of the myocardium by fatty or fibrofatty tissue, starting from the epicardium toward the endocardium and eventually becoming transmural.<sup>6</sup> The disease is familial in about half of the cases and typically shows an autosomal dominant inheritance, with reduced penetrance and variable expression.<sup>7</sup> A genetic cause can be established in about 60% of the patients.<sup>8</sup>

The majority of pathogenic variants in ACM have been identified in genes encoding the intercalated disc (ID) proteins plakoglobin (*JUP*)<sup>9</sup>, desmoplakin (*DSP*)<sup>10</sup>, plakophilin-2 (*PKP2*)<sup>11</sup>, desmoglein-2 (*DSG2*)<sup>12</sup>, desmocollin-2 (*DSC2*)<sup>13</sup>, alpha-T-catenin (*CTNNA3*)<sup>14</sup> and N-cadherin (*CDH2*).<sup>15,16</sup>

A few other mutations have been found in non-ID genes encoding cardiac ryanodine receptor 2 (*RYR2*)<sup>17</sup>, transforming growth factor beta 3 (*TGFβ3*)<sup>18</sup>, transmembrane protein 43 (*TMEM43*)<sup>19</sup>, desmin (*DES*)<sup>20</sup>, titin (*TTN*)<sup>21</sup>, lamin A/C (*LMNA*)<sup>22</sup>, phospholamban (*PLN*)<sup>23</sup> and sodium voltage-gated channel alpha subunit 5 (*SCN5A*).<sup>24</sup>

Here we report the identification of likely pathogenic variants in the tight junction protein 1 (*TJPI*) gene in patients with ACM, thereby linking pathogenic variants in *TJPI* with a human disease for the first time. *TJPI* encodes for a scaffolding protein, called zonula occludens-1 (ZO-1), which interacts with various components of cardiac ID.

## Materials and Methods

Raw data, analytic methods, and study materials are available on request from the corresponding author, for purposes of reproducing the results or replicating the procedure. The study protocol was approved by the ethics committee review board of the University of Padua (Padua, Italy) and of the Amsterdam University Medical Center (Amsterdam, The Netherlands).

Methods are available as supplemental data.

## **Results**

### **Genetic screening**

Proband II-4 of family A, negative for point mutations and copy number variations (CNVs) in the most frequently mutated ACM genes, was screened by whole exome sequencing (WES). After the removal of sequencing adapters and trimming of low-quality bases, approximately 25 Gb of cleaned sequencing bases were obtained. 99.99% of the sequencing reads was aligned to the human reference genome (hg19), with 63.17% effective reads from target regions being obtained after the removal of PCR duplications. The average sequencing depth was 216.57-fold, with 98.57% of target regions having at least a 10-fold coverage.

As a result, 21711 SNPs and 644 InDels were identified. Upon excluding synonymous variants outside splicing regulatory sites and predicted benign missense variants, 52 novel or rare variants with a  $MAF \leq 0.01\%$  were filtered (Table I in the Data Supplement). *TJP1* which encodes a multifunctional protein interacting with various components of the cardiac ID called ZO-1, resulted the first ranked gene by two different prioritization tools (Figure 1A and Table II in the Data Supplement). The p.(Y669C) missense variant was absent from all databases and was predicted to be damaging by 8 *in silico* tools (Figure 1B). This variant has a CADD C-score of 25.5 (Figure 1B); of note, a C-score over 20 indicates that it is amongst the top 1% of deleterious variants in the human genome. p.(Y669C) affects a highly conserved amino acid at

the guanylate kinase-like (GUK) domain of ZO-1 (Figure 1C). Sanger sequencing confirmed the presence of this heterozygous variant in the proband and in 6 out of 9 family members (Figure 2).

The presence of additional variants in exons with a coverage <15X of known ACM genes was excluded by Sanger sequencing in the proband's DNA.

Targeted genetic screening of *TJPI* in 40 additional Italian ACM patients identified the likely pathogenic variant c.793C>T p.(R265W) in the proband of family B. This variant is present in gnomAD database with a MAF of 0.001% (2/111550 alleles) in European individuals. It is predicted to be damaging by 8 *in silico* tools and its CADD C-score is 33 (Figure 1B), indicating that it is amongst the top 0.1% of deleterious variants in the human genome. This variant affects a highly conserved amino acid localized just after the end (at position 261) of the second PDZ domain of ZO-1 (Figure 1C).

To exclude the presence of additional mutations in ACM genes or in other candidate genes, the proband's DNA was also analyzed by WES. After removing low-quality reads, approximately 18.2 Gb of cleaned sequencing bases were obtained; 99.98% of the sequencing reads was aligned to the human reference genome (hg19), with 62.93% effective reads from target regions being obtained after the removal of PCR duplications. The average sequencing depth was 162.26-fold, with 98.07% of target regions having at least a 10-fold coverage. As a result, 21691 SNPs and 630 InDels were identified. Among them, 48 variants were filtered as described above (Table III in the Data Supplement). Once again, *TJPI* resulted the first ranked gene by two different prioritization tools (Figure 1A and Table IV in the Data Supplement). The presence of additional variants was excluded in all exons of ACM genes with a coverage <15X by Sanger sequencing. In family B, the affected proband's sister (II-3) turned out to also have this variant (Figure 2).

In addition, WES data has been analyzed from an independent cohort of 43 Dutch/German patients with a clinical diagnosis of ACM; among them, 11 fulfilled the formal criteria for its sub-form ARVC while 32 had a diagnosis of non-ischemic dilated cardiomyopathy (DCM) with recurrent sustained ventricular tachycardia (VT). Two missense pathogenic variants in *TJPI* (c.986C>T, p.(S329L) and c.1079A>T, p.(D360V)) have been identified in two probands. As before, *TJPI* gene was ranked first by Endeavour in both patients (Figure 1A and Tables VI and VII in Data Supplement). The p.(S329L) variant, identified in a proband negative for mutations in ACM genes, is present in gnomAD database with a MAF of 0.006% (16/269882 alleles), and is predicted to be damaging by 6 *in silico* tools and its CADD C-score is 30 (Figure 1B). The p.(D360V) variant, identified in a proband carrying also the p.(I156N) pathogenic variant in *DSP*, is absent from all databases, is predicted to be damaging by 5 *in silico* tools and its CADD C-score is 28.5 (Figure 1B). In the absence of additional affected family members, we cannot define the pathogenic role of *TJPI* and *DSP* variants.

Both *TJPI* variants affect highly conserved amino acids localized between PDZ2 and PDZ3 domains. Sanger sequencing confirmed the presence of these heterozygous variants in both probands.

Interestingly, *TJPI* showed a probability of being a loss-of-function (LoF) intolerant (pLI) gene of 1. A gene with a pLI value  $\geq 0.9$  is considered extremely LoF intolerant. In gnomAD database 1160 *TJPI* variants with a MAF  $\leq 0.01\%$  were reported in 1415 out of 138632 (1%) subjects. Among these variants, only 228 were predicted to have a probably-damaging effect and were identified in 272 out of 138632 (0.19%) subjects. On the other hand, we identified 4 likely pathogenic variants in 4 out of 84 (4.8%) ACM patients, hence showing a significantly higher prevalence compared to individuals carrying a rare variant ( $p=0,0006$ ) and to individuals carrying a probably-damaging variant ( $p=6,1354E-21$ ).



## Clinical Data

Family A was evaluated after the sudden death of a 36 years old man (II-2); a few months before, he presented with a syncopal episode during effort. ECG showed negative T waves in V1-V3, whereas on a 2D-echocardiogram the right ventricle (RV) was dilated (PLAX RVOT = 42 mm; PSAX RVOT = 41 mm) with a reduced systolic function (FAC: 30%) and presence of kinetic abnormalities (Table 1 and Table VIII in Data Supplement). Before a cardiac magnetic resonance (MRI) was made, he died during effort. Autopsy was performed at another hospital leading to a diagnosis of an unspecified cardiomyopathy. Histologic slides, which were later reviewed at our center, showed fibro-fatty replacement in the free wall of both ventricles together with myocardial atrophy, features in keeping with ACM (Figure 3). The 35 years-old brother (II-4) was then evaluated in our outpatient clinic and diagnosed with ACM (Table 1). ECG showed intraventricular conduction delay, negative T waves in V1-V5 and low QRS voltages (Figure 4A). At 24-hour Holter ECG, frequent premature ventricular contractions (PVCs) with left bundle branch block (LBBB) morphology were detected. MRI showed RV dilatation with regional kinetic abnormalities; the left ventricle was also dilated with mild diffuse hypokinesia. Moreover, late gadolinium enhancement as mid-mural striae was detectable in the inferior-lateral wall of left ventricle (Figure 4B). One additional brother (II-9) fulfilled the ACM diagnostic criteria. The proband's mother (I-1) was asymptomatic and showed negative T waves in V1-V2, whereas the 2D echo was normal. Finally, subjects III-1, III-2 and III-5, who were children at the time of clinical examination, and the remaining proband's brother (subject II-6) did not show clinical and instrumental signs of the disease, in line with ACM incomplete penetrance (Table 1).

The proband of family B was a female diagnosed with ACM at the age of 51. She underwent cardiac evaluation for palpitations. ECG showed negative T waves in V1-V3 and frequent PVCs at 24-hour Holter ECG. 2D-echo revealed a moderated dilatation and dysfunction of the RV

(Table 1). During the familial screening, the 65 years-old sister was also diagnosed with ACM, showing negative T waves in V1-V2 and frequent PVCs with LBBB-morphology. Her 2D-echo showed mild RV dilation with normal function (Table 1).

Proband C carrying the p.(S329L) variant was a male with DCM at the age of 55 years. He underwent cardiac evaluation due to progressive dyspnea. His LV ejection fraction was estimated at 33%, with a LVED = 76 mm. RV function was considered normal, with just moderately depressed contractility. The family history was negative. ECG showed no repolarization abnormalities. He was further known with persistent atrial fibrillation, renal dysfunction, diabetes and hypertension. At 61 he had an electrical storm that was treated by medication. At the age of 62 years he was referred to the hospital for VT ablation due to recurrent VTs with ICD shocks. Electro-anatomical mapping revealed an arrhythmogenic substrate in the posterior-basal LV epicardium and in the anterior-basal septal endocardium. He had VT recurrence after one year. Six months later he received a LVAD due to progression of heart failure.

Proband D, carrying the p.(D360V) variant, was a male with DCM diagnosed at the age of 66. He underwent cardiac evaluation due to syncope and documentation of fast VTs with right bundle branch block-morphology. LV ejection fraction was estimated at 40% (LVEDD = 67 mm). RV function was considered normal. The family history was negative. ECG showed no repolarization abnormalities. He was further known with diabetes and hypertension. At 73 he had an electrical storm and was referred to the hospital for VT ablation. Endocardial electro-anatomical mapping revealed an arrhythmogenic substrate in the inferior-posterior basal LV with a septal extension. At the 2-year follow-up he had no VT recurrence.

### ***In silico* prediction of the effects of ZO-1 likely pathogenic variants**

ZO-1 is a multi-domain protein, presenting three consecutive PDZ (PDZ1-3) domains, a SH3 domain, a guanylate kinase-like (GUK) domain and a ZU5 domain, as reported in the PFAM database. Our alignment of 43 orthologous sequences showed that amino acid R265 and the 9 residues immediately downstream are fully conserved between species. We then checked if the structural characteristics of the ZO-1 fragment <sup>265</sup>RATLLNVPDL<sup>275</sup> could be impaired by the pathogenic variant. Predictions of local structure content with FIELDS of both wild type and mutant suggested that this fragment forms an additional PDZ2  $\beta$ -strand, whereas the p.(R265W) variant induces destabilization of this transient strand (Figure 5A).

Both p.(S329L) and p.(D360V) variants localize in an intrinsically disordered linker connecting PDZ2 and PDZ3. Multiple sequence alignment shows that the entire region, including the two positions, is almost conserved among mammals. The p.(S329L) variant overlaps five putative linear motifs (Figure S1 in the Data Supplement); three of them suggest that this region may be involved in protein-protein interaction. Similarly, p.(D360V) variant is predicted to overlap three linear motifs (Figure S1 in the Data Supplement), two of them suggest that this mutation may impair ZO-1 interaction with MAP kinases. However, the prediction confidence is low, thus not allowing further speculation about its pathogenic effect.

The p.(Y669C) likely pathogenic variant is localized within the GUK domain. This enzymatically inactive domain is connected to the PDZ3 domain through a SH3 module acting as a linker.<sup>25</sup> The variant falls at the contact region between strands  $\beta$ 2/ $\beta$ 3, which are members of a large five  $\beta$ -sheet structure which forms the GUK hydrophobic core. We performed a 40 ns full atom molecular dynamics simulation to investigate the structural impact of the p.(Y669C) variant. Analysis of the trajectories showed the wild-type protein to be relatively more stable than the mutant. In particular, root mean square deviation (RMSD) analysis showed the wild type protein fluctuating around 0.3 nm (Figure 5B). We also observed a short peak of 0.5 nm between 12 and 18 ns of simulation. Frames inspection revealed that the increase was

due to the movement of a single loop within residues 586-631. In the same simulation conditions, the mutant protein showed a RMSD peak of 0.5 nm, reached after only few ns, while the RMSD value remains over 0.4 nm for the entire simulation. Structural comparison between the last frame calculated for each run showed that p.(Y669C) could induce a visible structural rearrangement of the GUK domain. When superimposed, the proteins showed a similar organization for the PDZ3 domain, while the GUK domain resulted to adopt a more opened conformation in the mutated protein (Figure 5B). Hence, we propose that increased mobility of the GUK domain is compatible with a local unfolding induced by p.(Y669C). A network of interacting residues generated with RING around 10 Å from position 669 showed that several hydrophobic interactions were lost during the simulation of the mutant protein. In particular, p.(Y669C) broke the van der Waals interactions between Y669 and L66, I789, promoting a clear structural rearrangement.

The crystal structure of the ZO-1 N-terminal domain, in which p.(R265W), p.(S329L) and p.(D360L) variants are located, is not available. Therefore, a molecular dynamics simulation could not be carried out.

## **Discussion**

Here we provide the first description of likely pathogenic *TJPI* variants identified in patients with ACM; no previous association of *TJPI* variants to inherited human diseases has been reported. The four probands fulfil the current diagnostic criteria for ARVC or had DCM with recurrent sustained VTs and three of them resulted negative for pathogenic/likely pathogenic variants in all known ACM genes and for CNVs in the most frequently mutated genes. Moreover, no additional pathogenic variants have been identified in obvious ACM candidate genes by WES analyses. Mutation screening performed in all available family members revealed that all affected subjects were mutation carriers.

*TJP1* gene encodes the multifunctional protein ZO-1, which comprises four different domains, PDZ, SH3, GUK and ZU5, and interacts with different proteins of the ID.

This latter is a specialized “organelle” in which mechanical and electrical junctions are interlinked via protein-protein interactions.<sup>26</sup> The ID components were initially described as independent; however, it has been demonstrated that in cardiac tissue of higher order mammalian species they actively interact and form multiprotein complexes called *area composita* (constituted by proteins of desmosomes and adherens junctions) and *connexome* (constituted by proteins of desmosomes, gap junctions and sodium channel complex).<sup>26</sup>

In cardiomyocytes, ZO-1 is essential for a normal organization of both gap junctions (GJs) and *area composita*, interacting with Connexin43 (Cx43) and N-Cadherin respectively.<sup>27,28</sup> It has been reported that ZO-1 regulates the number, size and distribution of GJs, binding Cx43 molecules present at the *perinexus*, an area adjacent to the functional GJs.<sup>29,30</sup> Free connexons in the *perinexus* are unable to associate with connexons of neighboring cells as long as ZO-1 is bound.<sup>26</sup> In ventricles from patients with congestive heart failure, showing GJs remodeling, ZO-1 protein levels were significantly increased, whereas Cx43 expression was downregulated; the proportion of Cx43 interacting with ZO-1 was significantly enhanced, thus suggesting that ZO-1 prevents the formation of functional GJs.<sup>27</sup> Interestingly, Palatinus et al. reported that the regulation of functional GJ localization at cell-cell interactions between cardiomyocytes does not depend on the ZO-1-Cx43 interaction only, but also on the level of ZO-1 association with the N-cadherin in the *area composita*.<sup>28</sup> ZO-1 forms a molecular link between GJs and *area composita*, interacting with Cx43 via the second PDZ domain<sup>31</sup> and connecting cadherin to the actin filaments through the interaction with  $\alpha$  catenin.<sup>32,33</sup>

Tjp1 knockout mice showed embryonic lethal phenotype associated with massive apoptosis in the notochord, neural tube area and allantois, as well as defective yolk sac angiogenesis, hence indicating an essential role of Tjp1 in early embryonic development. Heterozygous Tjp1<sup>+/-</sup> mice

showed no obvious phenotypes, with fertility and growth rates comparable to those of wild-type mice.<sup>34</sup> Moreover, established ZO-1 knockout Madin-Darby canine kidney cells showed remarkable changes comparable in myosin organization at cell-cell contacts and an altered localization of tight junction proteins. These changes were reversed by trace ZO-1 expression, although an excessive ZO-1 level induced an altered shape of cell-cell junctions.<sup>35</sup>

The four likely pathogenic missense ZO-1 variants reported in the present study affect highly conserved amino acid residues and are predicted as deleterious by at least 5 prediction tools, hence suggesting a functional role for these residues. The p.(R265W) pathogenic variant is located in a conserved extension of the PDZ2 domain. Due to its localization, formally outside the PDZ2, this variant could be considered not relevant for correct PDZ domain folding. Previous structural investigation conducted through nuclear magnetic resonance spectroscopy has indeed shown that this residue falls in an intrinsically disordered region.<sup>36</sup> However, it has been described that disordered regions may undergo specific phase transitions depending on their amino acid composition and/or binding partners.<sup>37</sup> Our predictions of the local structure content of both wild type and mutant proteins suggest that the intrinsically disordered fragment <sup>265</sup>RATLLNVPDL<sup>275</sup> forms an additional PDZ2  $\beta$ -strand, whereas p.(R265W) variant induces its destabilization. Several PDZ domains are known to present extra regulative regions that extend over the C-terminus boundary and frequently assume a regulative role, e.g. improving domain stability and/or modulating the ligand binding specificity.<sup>38,39</sup> As a relevant example, PDZ3 of ZO-1 presents itself an accessory C-terminal tail, which acts as both a stabilizer of PDZ3/SH3 domain-domain interaction and a ligand binding site.<sup>40</sup> Due to the high conservation and the results of our *in silico* investigation, we hypothesize that the p.(R265W) variant could impair the regulative function of the PDZ2-C-terminus region.

The other two likely pathogenic variants p.(S329L) and p.(D360V) localize in a linker region connecting PDZ2 and PDZ3 and overlap different putative linear motifs involved in protein-

protein interaction. Large-scale studies of human phosphor-proteome showed that the S329 is a phosphorylation site thought to be implicated in embryotic differentiation and cell mitosis.<sup>41,42</sup> One may speculate that p.(S329L) impairs ZO-1 association with an unknown protein containing a WW-domain, a modular protein domain that binds to phosphoserine containing motifs.<sup>43</sup> However, the pathogenic effect of these two variants may be mostly related to a local destabilization of the disordered linker, rather than to a specific functional impairment.

The p.(Y669C) variant falls inside the GUK domain, for which a crystal structure is available.<sup>25</sup> Molecular dynamics simulation revealed that the mutant protein is less stable than the wild-type, showing a remarkable loss of hydrophobic interactions around the mutated area and a more open conformation of the GUK domain. Lye and coworkers presented a structural comparative study between ZO-1 and ZO-3, demonstrating that the main difference between the two proteins is a more open conformation of the GUK domain in ZO-3.<sup>25</sup> This variation in conformation between ZO-1 and ZO-3 might explain the inability of ZO-3 to localize to tight junctions in the absence of either ZO-1 or ZO-2. Our findings suggest that the local unfolding induced by the p.(Y669C) variant promotes a structural rearrangement which forces the ZO-1 GUK domain to assume a conformation closer to ZO-3.

Taking into account the localization of the ZO-1 likely pathogenic variants and the role of ZO-1 in the cardiac ID, we hypothesize that the mutated proteins alter their binding with one or more of the main ZO-1 interacting proteins, such as Cx43 or N-cadherin, thus causing a possible impairment of the *area composita* and/or the *connexome*. In this perspective, the identification of ZO-1 pathogenic variants in ACM patients underlines the potential direct involvement in ACM pathogenesis not only of the mechanical components of the ID (i.e. *area composita* and *desmosome*) but also of the electrical ones (gap junction).

In summary, multiple lines of evidence, such as cardiac expression of *TJP1* and function of the encoded protein ZO-1, co-segregation of the variants with the phenotype in two families, high

conservation of involved amino acids and the *in silico* predictions of pathogenicity and of the effects of the variants, suggest a causal relationship between *TJPI* pathogenic variants and ACM. Therefore, since the prevalence of *TJPI* variants in our cohort of ACM patients is just below 5%, we propose a systematic screening for this gene to determine the prevalence of variants in different cohorts of ACM patients in a research setting. Further studies will be needed to better understand the pathogenic effects of these variants, the behavior of the mutant ZO-1 proteins *in vivo* and the molecular mechanism leading to the ACM phenotype.

### **Acknowledgments**

We are grateful to Dr. Raffaele Lopreiato, Dr. Geppo Sartori, Department of Biomedical Sciences, University of Padua; Prof. Elisa Greggio, Dr. Libero Vitiello, Department of Biology, University of Padua for advice provided and for critical and helpful discussions on the manuscript. Rob Zwart is acknowledged for his help in analyzing Dutch and German samples. We also wish to thank the patients and their families for participating to the study.

### **Funding Sources**

This study was funded by the TRANSAC Strategic Research Grant CPDA133979/13, University of Padua, Italy; Target Projects 331/12, RP 2014-00000394, Regional Health System, Venice, Italy; The University of Padua Research Project (PRAT) CPDA133979; Registry for Cardio-cerebro-vascular Pathology, Veneto Region, Venice, Italy; BIRD162733, University of Padua, Italy; PRIN Ministry of Education, University and Research Grant 2015ZLNETW, Rome, Italy.

We wish to acknowledge funding from the Netherlands Cardiovascular Research Initiative, an initiative supported by the Dutch Heart Foundation (CVON2012-10 PREDICT and CVON2015-12 eDETECT).



## Disclosures

None.

## References

1. Corrado D, et al. Trends in sudden cardiovascular death in young competitive athletes after implementation of a preparticipation screening program. *JAMA*. 2006;296:1593-1601.
2. Thiene G, et al. Right ventricular cardiomyopathy and sudden death in young people. *N Engl J Med*. 1988;318:129-133.
3. Nava A, et al. Familial occurrence of right ventricular dysplasia: a study involving nine families. *J Am Coll Cardiol*. 1988;12:1222-1228.
4. Rastegar N, et al. Spectrum of biventricular involvement on CMR among carriers of ARVD/C-associated mutations. *JACC Cardiovasc Imaging*. 2015;8:863-864.
5. Sen-Chowdhry S, et al. Left-dominant arrhythmogenic cardiomyopathy: an under-recognized clinical entity. *J Am Coll Cardiol*. 2008;52:2175-2187.
6. Basso C, et al. Arrhythmogenic right ventricular cardiomyopathy. Dysplasia, dystrophy, or myocarditis? *Circulation*. 1996;94:983-991.
7. Nava A, et al. Clinical profile and long-term follow-up of 37 families with arrhythmogenic right ventricular cardiomyopathy. *J Am Coll Cardiol*. 2000;36:2226-2233.

8. Poloni G, et al. Arrhythmogenic right-ventricular cardiomyopathy: molecular genetics into clinical practice in the era of next generation sequencing. *J Cardiovasc Med (Hagerstown)*. 2016;17:399-407.
9. McKoy G, et al. Identification of a deletion in plakoglobin in arrhythmogenic right ventricular cardiomyopathy with palmoplantar keratoderma and woolly hair (Naxos disease). *Lancet*. 2000;355:2119-2124.
10. Rampazzo A, et al. Mutation in human desmoplakin domain binding to plakoglobin causes a dominant form of arrhythmogenic right ventricular cardiomyopathy. *Am J Hum Genet*. 2002;71:1200-1206.
11. Gerull B, et al. Mutations in the desmosomal protein plakophilin-2 are common in arrhythmogenic right ventricular cardiomyopathy. *Nat Genet*. 2004;36:1162-1164.
12. Pilichou K, et al. Mutations in desmoglein-2 gene are associated with arrhythmogenic right ventricular cardiomyopathy. *Circulation*. 2006;113:1171-1179.
13. Syrris P, et al. Arrhythmogenic right ventricular dysplasia/cardiomyopathy associated with mutations in the desmosomal gene desmocollin-2. *Am J Hum Genet*. 2006;79:978-984.
14. van Hengel J, et al. Mutations in the area composita protein alpha T-catenin are associated with arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J*. 2012;34:201-210.

15. Mayosi BM, et al. Identification of Cadherin 2 (CDH2) Mutations in Arrhythmogenic Right Ventricular Cardiomyopathy. *Circ Cardiovasc Genet*. 2017;10:e001605.
16. Turkowski KL, et al. Whole exome sequencing with genomic triangulation implicates CDH2-encoded N-cadherin as a novel pathogenic substrate for arrhythmogenic cardiomyopathy. *Congenit Heart Dis*. 2017;12:226-235.
17. Tiso N, et al. Identification of mutations in the cardiac ryanodine receptor gene in families affected with arrhythmogenic right ventricular cardiomyopathy type 2 (ARVD2). *Hum Mol Genet*. 2001;10:189-194.
18. Beffagna G, et al. Regulatory mutations in transforming growth factor-beta 3 gene cause arrhythmogenic right ventricular cardiomyopathy type 1. *Cardiovasc Res*. 2005;65:366-373.
19. Merner ND, et al. Arrhythmogenic right ventricular cardiomyopathy type 5 is a fully penetrant, lethal arrhythmic disorder caused by a missense mutation in the TMEM43 gene. *Am J Hum Genet*. 2008;82:809-821.
20. van Tintelen JP, et al. Severe cardiac phenotype with right ventricular predominance in a large cohort of patients with a single missense mutation in the DES gene. *Heart Rhythm*. 2009;6:1574-1583.
21. Taylor M, et al. Genetic variation in titin in arrhythmogenic right ventricular cardiomyopathy-overlap syndromes. *Circulation*. 2011;124:876-885.

22. Quarta G, et al. Mutations in the Lamin A/C gene mimic arrhythmogenic right ventricular cardiomyopathy. *Eur Heart J*. 2012;33:1128-1136.
23. van der Zwaag PA, et al. Phospholamban R14del mutation in patients diagnosed with dilated cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy: evidence supporting the concept of arrhythmogenic cardiomyopathy. *Eur J Heart Fail*. 2012;14:1199-1207.
24. Te Riele AS, et al. Multilevel analyses of SCN5A mutations in arrhythmogenic right ventricular dysplasia/cardiomyopathy suggest non-canonical mechanisms for disease pathogenesis. *Cardiovasc Res*. 2017;113:102-111.
25. Lye MF, et al. Insights into regulated ligand binding sites from the structure of ZO-1 Src homology 3-guanylate kinase module. *J Biol Chem*. 2010;285:13907-13917.
26. Vermij SH, et al. Refining the molecular organization of the cardiac intercalated disc. *Cardiovas Res*. 2017;113:259-275.
27. Bruce AF, et al. Gap junction remodeling in human heart failure is associated with increased interaction of connexin43 with ZO-1. *Cardiovasc Res*. 2008;77:757-765.
28. Palatinus JA, et al. ZO-1 determines adherens and gap junction localization at intercalated disks. *Am J Physiol Heart Circ Physiol*. 2011;300:583-594.

29. Rhatt JM, et al. The perinexus: A new feature of Cx43 gap junction organization. *Heart Rhythm*. 2012;9:619-623.
30. Rhatt JM, et al. Connexin 43 connexon to gap junction transition is regulated by zonula occludens-1. *Mol Biol Cell*. 2011;22:1516-1528.
31. Giepmans BN, et al. The gap junction protein connexin43 interacts with the second PDZ domain of the zona occludens-1 protein. *Curr Biol*. 1998;8:931-934.
32. Itoh M, et al. Involvement of ZO-1 in cadherin-based cell adhesion through its direct binding to alpha catenin and actin filaments. *J Cell Biol*. 1997;138:181-192.
33. Yokoyama S, et al. alpha-catenin-independent recruitment of ZO-1 to nectin-based cell-cell adhesion sites through afadin. *Mol Biol Cell*. 2001;12:1595-609.
34. Katsuno T, et al. Deficiency of zonula occludens-1 causes embryonic lethal phenotype associated with defected yolk sac angiogenesis and apoptosis of embryonic cells. *Mol Biol Cell*. 2008;19:2465-2475.
35. Tokuda S, et al. ZO-1 knockout by TALEN-mediated gene targeting in MDCK cells: involvement of ZO-1 in the regulation of cytoskeleton and cell shape. *PLoS One*. 2014;9:e104994.
36. Ji P, et al. Solution structure of the second PDZ domain of Zonula Occludens 1. *Proteins*. 2011;79:1342-1346.

37. Wu H, et al. The Structure and Dynamics of Higher-Order Assemblies: Amyloids, Signalosomes, and Granules. *Cell*. 2016;165:1055-1066.
38. Wang CK, et al. Extensions of PDZ domains as important structural and functional elements. *Protein Cell*. 2010;1:737-751.
39. Zhang J, et al. Phosphorylation of a PDZ domain extension modulates binding affinity and interdomain interactions in postsynaptic density-95 (PSD-95) protein, a membrane-associated guanylate kinase (MAGUK). *J. Biol. Chem*. 2011;286:41776-41785.
40. Nomme J, et al. The Src homology 3 domain is required for junctional adhesion molecule binding to the third PDZ domain of the scaffolding protein ZO-1. *J. Biol. Chem*. 2011;286:43352-43360.
41. Olsen JV, et al. Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci Signal*. 2010;3:ra3.
42. Rigbolt KTG, et al. System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci Signal*. 2011;4:rs3.
43. Chen HI, et al. The WW domain of Yes-associated protein binds a proline-rich ligand that differs from the consensus established for Src homology 3-binding modules. *Proc Natl Acad Sci U S A*. 1995;92:7819-7823.



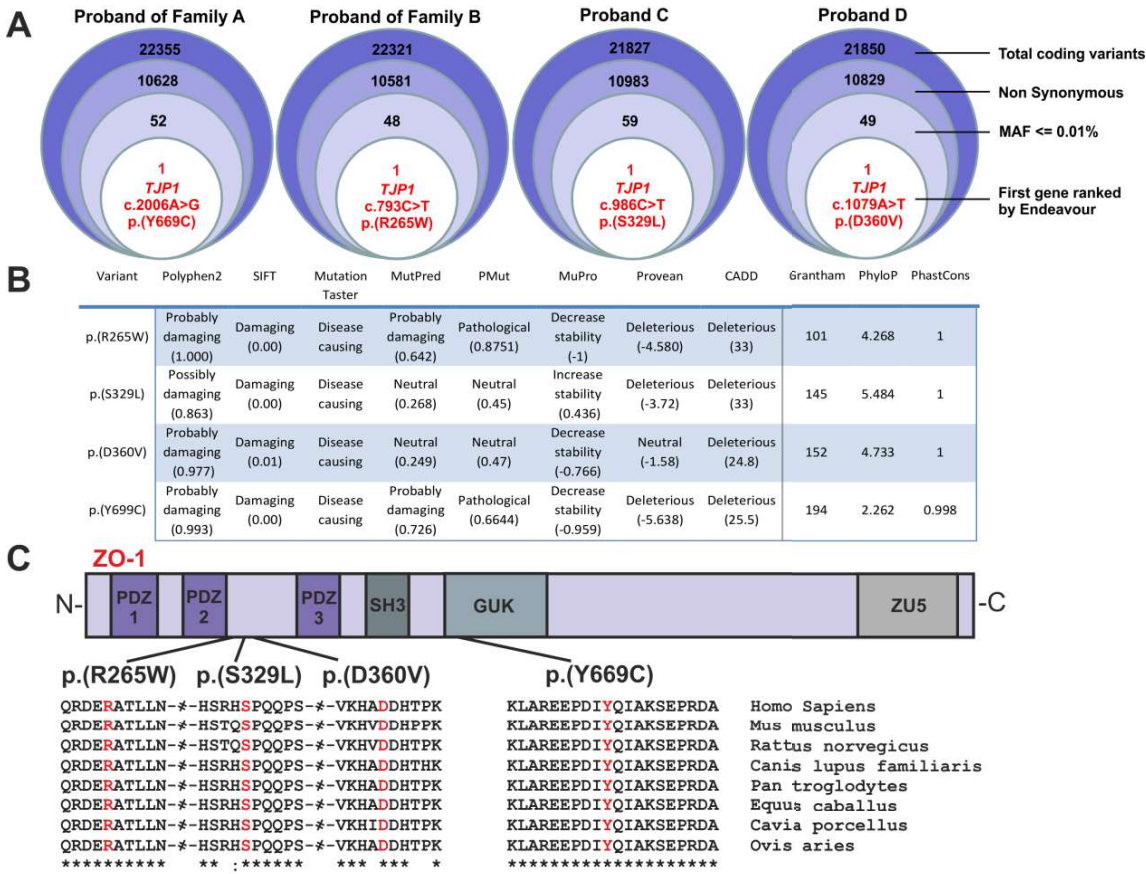
**Table 1.** Clinical data of ARVC probands and family members carrying the *TJPI* p.(R265W) and p.(Y669C) likely pathogenic variants

Family Subject	Gender	Age (yrs) at diagnosis/ last investigation	Family		Echocardiography RV global or regional dysfunction and structural alterations		Twelve-lead ECG repolarization abnormalities		Twelve-lead ECG depolarization/ conduction abnormalities		Arrhythmias	LV involved	ARVC diagnostic criteria		
			M	m	M	m	M	m	M	m					
Fam.A I-1	F	68	1	0	0	0	0	1	0	0	0	0	0	1M, 1m	
Fam.A II-2	M	36	0	0	1	0	1	0	0	0	NP	1	0	2M, 1m	
Fam.A II-4	M	35	1	0	1	0	1	0	0	0	1	0	1	3M, 2m	
Fam.A II-6	M	45	1	0	0	0	0	0	0	0	0	0	0	1M	
Fam.A II-9	M	40	1	0	0	0	0	0	0	0	1	0	1	0	1M, 2 m
Fam.A III-1	M	10	1	0	0	0	0	0	0	0	0	0	0	0	1M
Fam.A III-2	M	10	1	0	0	0	0	0	0	0	0	0	0	0	1M
Fam.A III-5	F	11	1	0	0	0	0	0	0	0	0	0	0	0	1M
Fam.B II-1	F	51	0	0	1	0	0	1	0	0	0	1	0	0	2M, 2m
Fam.B II-3	F	65	1	0	0	0	0	0	1	0	0	0	1	0	1M, 2m

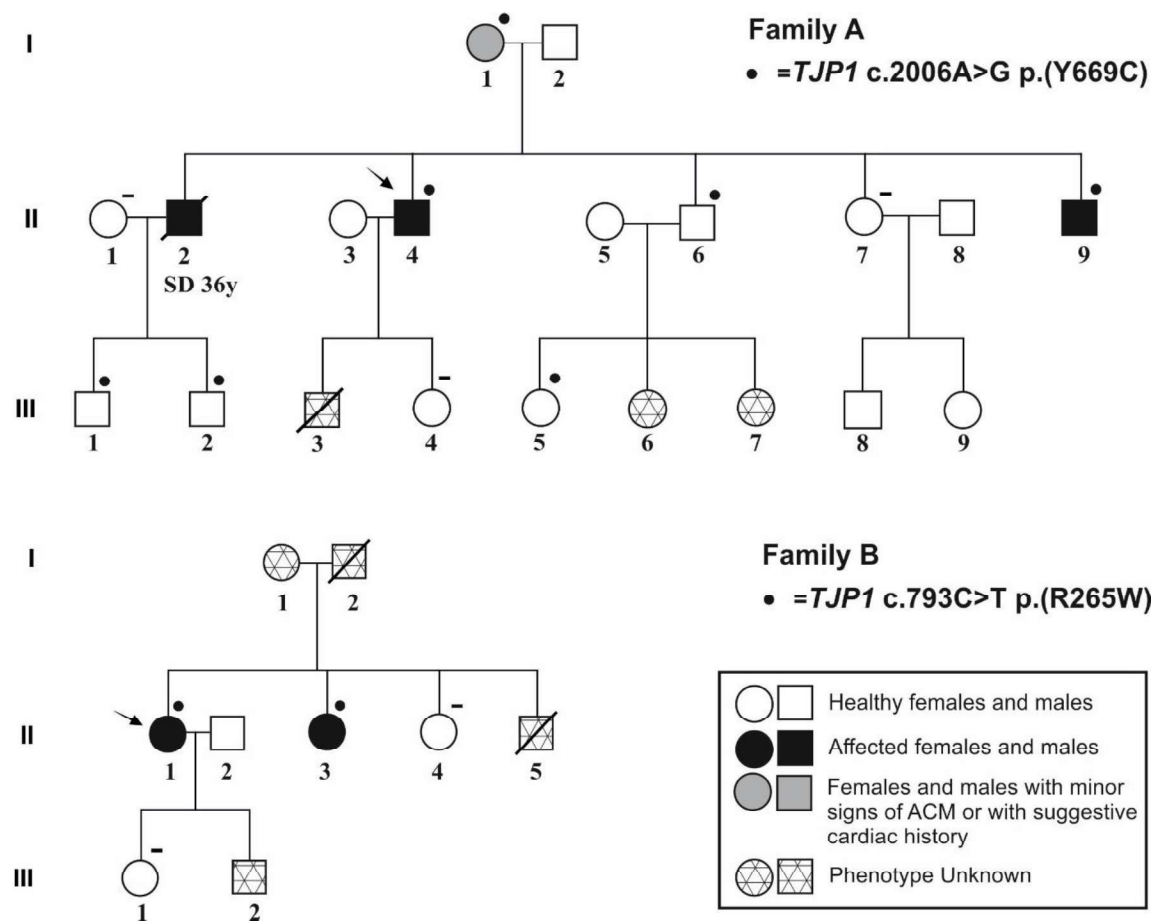
M: major criterion; m: minor criterion; 1: present; 0: absent.



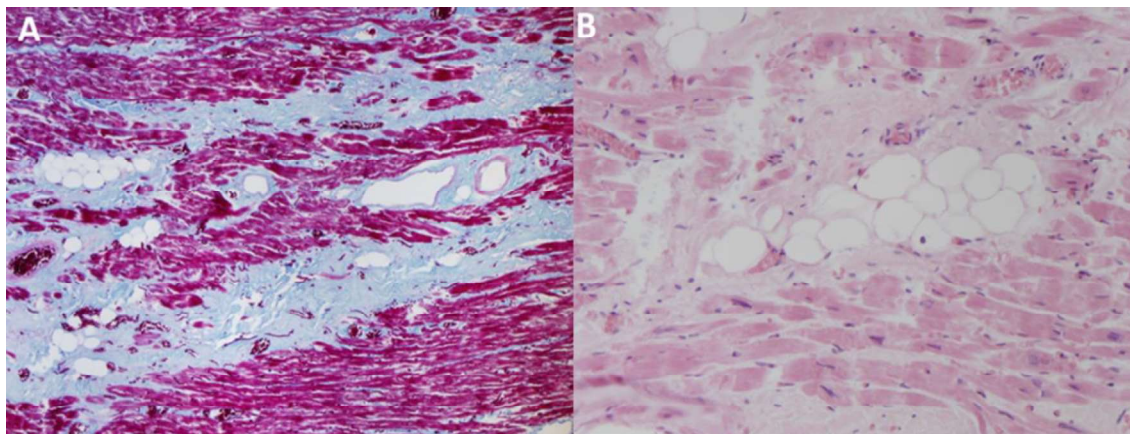
Figures and Figure Legends:



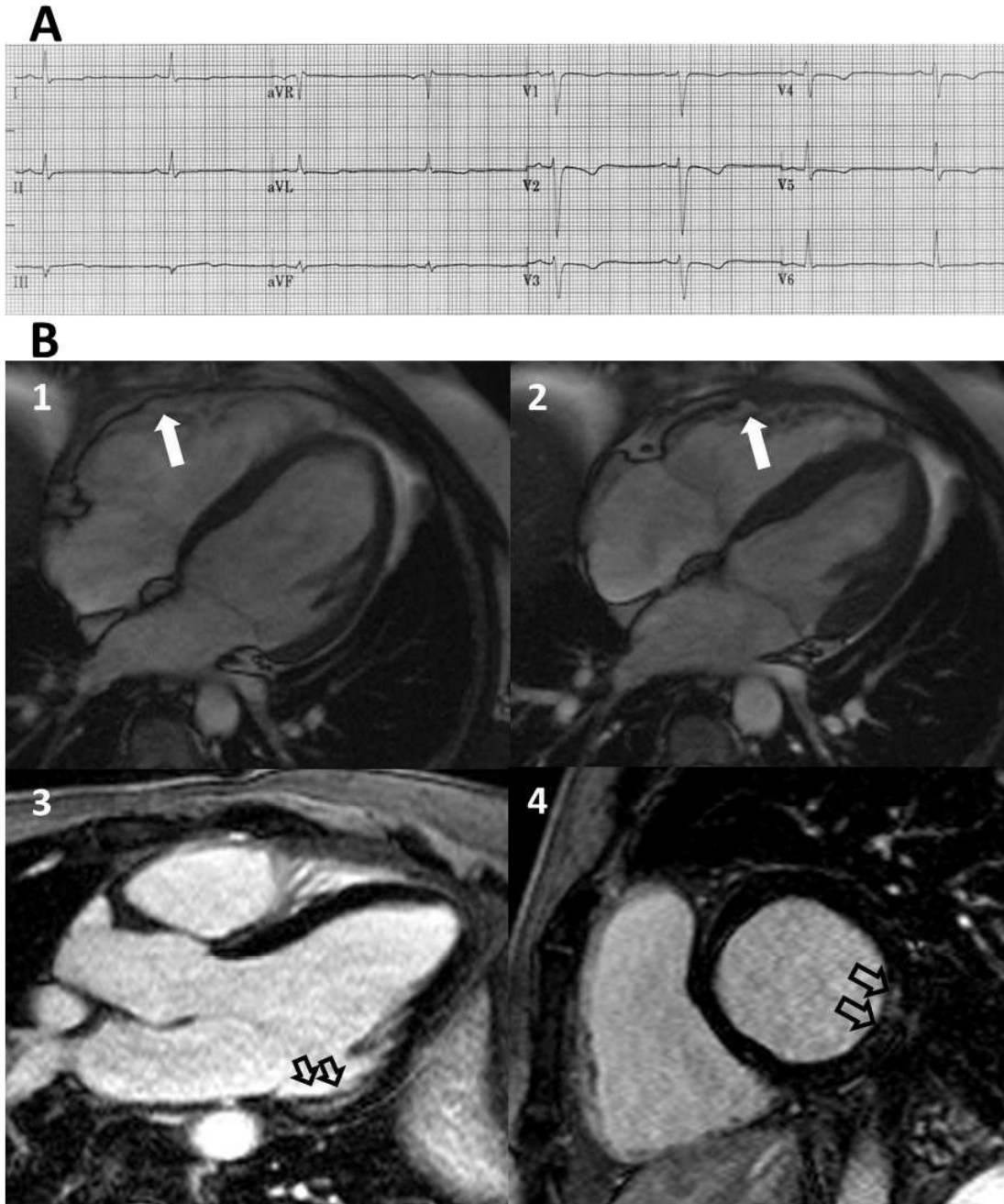
**Figure 1. *TJP1* likely pathogenic variants.** A) Schematic representation of variant filtering and gene prioritization performed in WES data of the four probands. B) *In silico* functional prediction and conservation scores of *TJP1* variants. C) Schematic representation of ZO-1 protein and localization of identified variants. Multi sequencing alignments showed evolutionary conservation of affected residues (highlighted in red) among species.



**Figure 2. Family pedigrees of patients carrying *TJP1* p.(Y669C) and p.(R265W) likely pathogenic variants. Arrows indicate the ACM probands.**

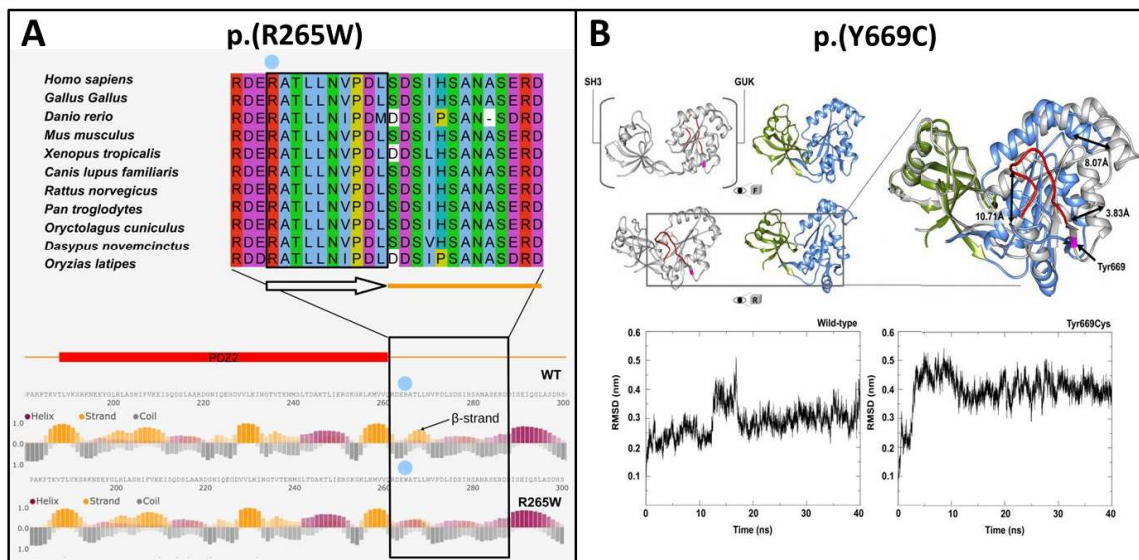


**Figure 3. Histological findings in the autoptic sample of patient II-2 of family A.** Histologic slide of the left ventricular free wall showing focal sub-epicardial fibro-fatty replacement in the setting of preserved wall thickness (A, trichrome staining; B, haematoxylin-eosin staining).



**Figure 4. Instrumental findings in the 35-year-old proband of family A carrying p.(Y669) likely pathogenic variant.** A) Twelve lead ECG shows sinus rhythm, intraventricular conduction delay, negative T waves in V1-V5 and low QRS voltages in limb leads. B) Cardiac magnetic resonance demonstrated a dilated right ventricle (end-diastolic volume 136 ml/m<sup>2</sup>) with impaired systolic function and focal bulging on the lateral wall (white arrow on diastolic frame in image 1 and on systolic frame in image 2). On T1-inversion recovery post-contrast

sequence late gadolinium enhancement as midmural striae was detectable in the inferior-lateral wall of left ventricle on 3-chamber long axis view (3), confirmed in the orthogonal short axis view (4) (empty arrows), suggesting a typical finding of arrhythmogenic biventricular cardiomyopathy.



**Figure 5. *In silico* predictions of effects of ZO-1 variants p.(R265W) and p.(Y669C).** A) Overview of the putative additional PDZ2 tail. (Top) Multiple sequence alignment of ZO-1 PDZ2 domain C-terminal flanking region. The extra boundary conserved region is marked with a black box, while a blue dot is used to mark the residue corresponding to the human R265. (Bottom) Latent local structure predictions from FIELDS suggest destabilization of the additional PDZ2  $\beta$ -strand. B) Overview and dynamics of ZO-1 SH3-GUK domains. (Top) Wild-type protein is presented in grey, while colored structure represents the p.(Y669C) mutant (front and rear views), with green and blue for SH3 and GUK domains respectively. Magenta represents the position of Y669, while red is used for the flexible loop within 586-631 residues. The superimposition of the last frames (40 ns) obtained after two independent runs of molecular dynamics simulation is shown on the right. Relative distances are presented as black arrows. Comparative view of the same structures shows a remarkable spatial reorganization of the region surrounding the position 669. (Bottom) Weighted RMSD fluctuation for ZO-1 region corresponding to the SH3-GUK domains. The two panels represent the variation observed for the wild type and p.(Y669C) mutant proteins, respectively. A significant variation in fold stability was observed for the mutant protein, which appears less stable than wild-type.